

<b>1. Introduction of Business</b>	<b>2</b>
1.1 Stakeholders & Background	2
1.2 Business problem to be resolved	2
<b>2. Describe the data</b>	<b>2</b>
<b>3. Exploratory data analysis</b>	<b>3</b>
3.1 Data Wrangling	3
3.1.1 List of postal codes of Toronto	3
3.1.2 Geospatial Coordinates	3
3.1.3 Visualization of the focused boroughs on map	4
3.1.4 Category Table from Foursquare	4
3.2 Clustering	5
3.2.1 Clustering based on the first category (Category_1 column)	5
3.2.2 Conclusion of Venues' feature based on first level categories	8
3.2.3 Cluster Neighborhoods based on the second level of venues' categories	9
<b>4. Conclusion</b>	<b>15</b>

---

# 1. Introduction of Business

## 1.1 Stakeholders & Background

Our customer is a food and beverage company. They have two brands: one is a cafe chain, the other one an Italian restaurant chain. They are searching for opportunities to develop in Toronto.

There are several aspects to be investigated for an investment, such as price of housing, tax policy, accessibility, and neighborhood environment.

But in this project, we will only focus on one aspect: neighborhood environment. We aim at providing a deep insight of the neighborhood environment of **Toronto city**, in order to cluster these neighborhoods for different types of restaurant investment.

## 1.2 Business problem to be resolved

How to choose wisely the location for a restaurant, depending on type?

- A good location to establish a **restaurant** should have:
  - good flow of quality customers
  - few competitors

We expect to look into the liveliness of the commercial area, including existing food and beverage businesses, hotels, parks, personal services, etc.. We also look into evaluating the competition: what kind of businesses already exist, are they dominating the market?

We will cluster the venues to analyze which neighborhood is more interesting for our customer. Finally we will give a list of neighborhoods for our customer, so they can proceed with investigating the other aspects.

---

## 2. Describe the data

- **List of postal codes of Toronto** : This dataset listed all the boroughs and neighborhoods of Toronto city that we will investigate. In this project we will only focus on the below Borough: Downtown Toronto, East Toronto, West Toronto, Central Toronto.

[https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)

	Postcode	Borough	Neighbourhood
0	M1A	Not assigned	Not assigned
1	M2A	Not assigned	Not assigned
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Harbourfront

- **Geolocation of Coordinates** : This dataset offers latitude and longitude of all the neighborhoods. It will help us with visualizing all the targeted location on the map.  
[http://cocl.us/Geospatial\\_data](http://cocl.us/Geospatial_data)

	<b>Postcode</b>	<b>Borough</b>	<b>Neighbourhood</b>
0	M1B	Scarborough	Rouge,Malvern
1	M1C	Scarborough	Highland Creek,Rouge Hill,Port Union
2	M1E	Scarborough	Guildwood,Morningside,West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae

- **Venue data from Foursquare API** : These dataset should be sufficient for our analysis, as the Foursquare API contains variance detailed data that we can look into.

## 3. Exploratory data analysis

### 3.1 Data Wrangling

#### 3.1.1 List of postal codes of Toronto

We firstly create dataframe of Toronto's postal code by scraping the following Wikipedia page [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M). Then, we need to proceed the following step to clean up this data:

- Only process the cells that have an assigned borough. Ignore cells with a borough that is Not assigned.
- More than one neighborhood can exist in one postal code area. For example, M5A is listed twice and has two neighborhoods: Harbourfront and Regent Park. These two rows will be combined into one row with the neighborhoods separated with a comma as shown in row 11 in the above table.
- If a cell has a borough but a Not assigned neighborhood, then the neighborhood will be the same as the borough.

#### 3.1.2 Geospatial Coordinates

Create a Dataframe using the geographical coordinates of each postal code from [http://cocl.us/Geospatial\\_data](http://cocl.us/Geospatial_data). Then merge it with the data frame of postal codes, the merged table shows as below:

	<b>Postcode</b>	<b>Borough</b>	<b>Neighbourhood</b>	<b>Latitude</b>	<b>Longitude</b>
0	M1B	Scarborough	Rouge,Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood,Morningside,West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

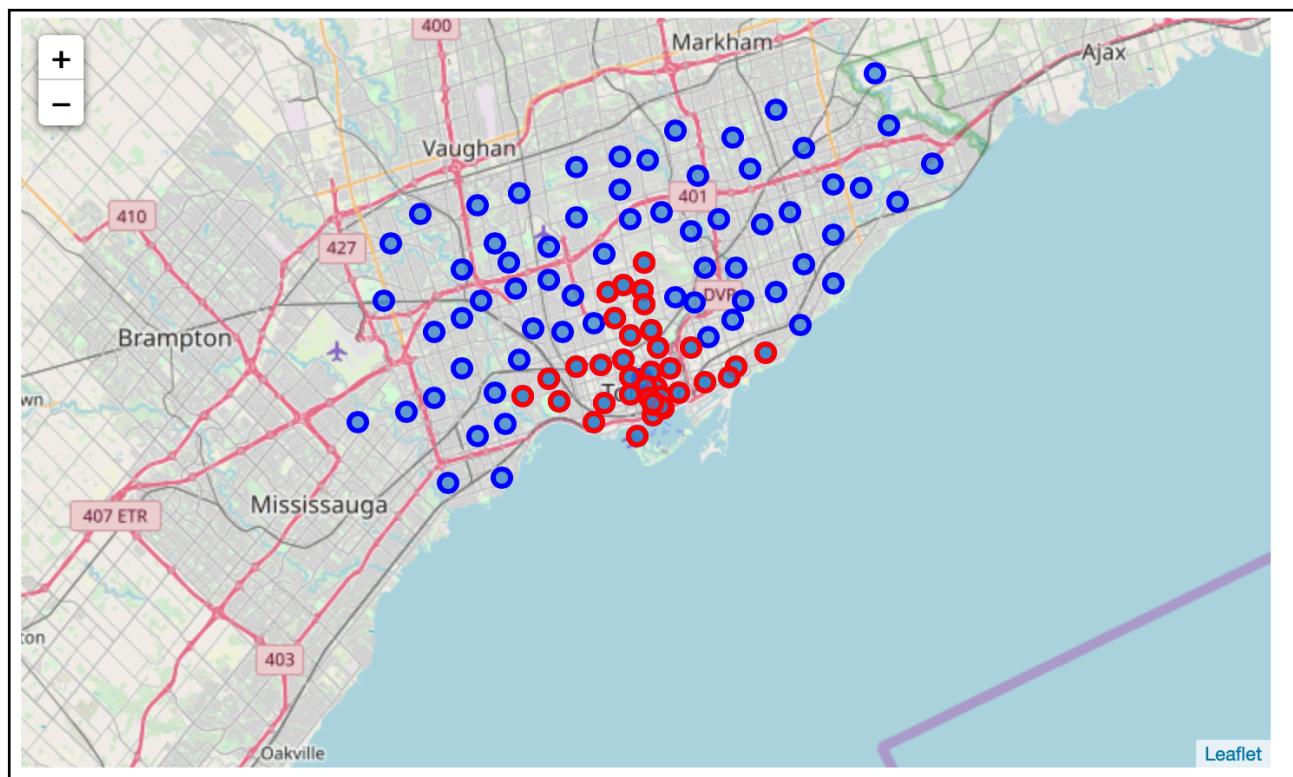
In this report, we will only focus on the below borough: Downtown Toronto, East Toronto, West Toronto, Central Toronto. So we will filter the above table and only look at the boroughs that we are interested:

Postcode	Borough	Neighbourhood	Latitude	Longitude
0	M4E	East Toronto	The Beaches	43.676357 -79.293031
1	M4K	East Toronto	The Danforth West,Riverdale	43.679557 -79.352188
2	M4L	East Toronto	The Beaches West,India Bazaar	43.668999 -79.315572
3	M4M	East Toronto	Studio District	43.659526 -79.340923
4	M4N	Central Toronto	Lawrence Park	43.728020 -79.388790

### 3.1.3 Visualization of the focused boroughs on map

From Python's library "Nominatim" we can get the latitude and longitude of Toronto. Using the library "folium", we can plot a map with neighborhoods superimposed on top of it.

The red points in the map below are the boroughs that we will work on.



### 3.1.4 Category Table from Foursquare

We can get information for all the types of venues from Foursquare's web site: <https://api.foursquare.com/v2/venues/categories>. But their data is not presented as a table. Therefore, we need to get the information structured, create a table of category.

There are in total 6 levels of categories. Element in Category\_1 are the topmost categories. There are 10 categories in the Category\_1 list:

Category_1	
0	Arts & Entertainment
1	College & University
2	Event
3	Food
4	Nightlife Spot
5	Outdoors & Recreation
6	Professional & Other Places
7	Residence
8	Shop & Service
9	Travel & Transport

Then, there is Category\_2, it includes more specified categories. Category\_3 is more specified, and so on. Here is the first 5 rows of the category table:

	Category_1	Category_2	Category_3	Category_4	Category_5	Category_6
0	Arts & Entertainment	Amphitheater	None	None	None	None
1	Arts & Entertainment	Aquarium	None	None	None	None
2	Arts & Entertainment	Arcade	None	None	None	None
3	Arts & Entertainment	Art Gallery	None	None	None	None
4	Arts & Entertainment	Bowling Alley	None	None	None	None

In this report, we will analyze two aspects for the location: the surrounding area and competition situation. We will use two type of categories to do so. Therefore this category table is very useful for the following analysis.

## 3.2 Clustering

### 3.2.1 Clustering based on the first category (Category\_1 column)

Our first goal is to analyze the surrounding environment, in order to know what types of businesses or facilities are in the neighbourhood.

#### 3.2.1.1 Get the Features

First, we get the top 100 venues from Foursquare API. We create a function named `getNearbyVenues`, to get the closest 100 venues, with respect to the neighbourhood center and within 500 meters.

	Borough	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	East Toronto	The Beaches	43.676357	-79.293031	Glen Manor Ravine	43.676821	-79.293942	Trail
1	East Toronto	The Beaches	43.676357	-79.293031	The Big Carrot Natural Food Market	43.678879	-79.297734	Health Food Store
2	East Toronto	The Beaches	43.676357	-79.293031	Grover Pub and Grub	43.679181	-79.297215	Pub
3	East Toronto	The Beaches	43.676357	-79.293031	Upper Beaches	43.680563	-79.292869	Neighborhood
4	East Toronto	The Beaches	43.676357	-79.293031	Dip 'n Sip	43.678897	-79.297745	Coffee Shop

We replace the column 'Venue Category' (which is too specific) in the above Data Frame by the corresponding first category (broader, and more relevant) in our category table in 3.4. For example, 'Trail' will be replaced by 'Outdoor & Recreation', 'Health Food Store' will be replaced by 'Shop & Service'.

After the replacement, we got the table below for the top 100 venues:

	Borough	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue new Category
0	East Toronto	The Beaches	43.676357	-79.293031	Glen Manor Ravine	43.676821	-79.293942	Outdoors & Recreation
1	East Toronto	The Beaches	43.676357	-79.293031	The Big Carrot Natural Food Market	43.678879	-79.297734	Shop & Service
2	East Toronto	The Beaches	43.676357	-79.293031	Grover Pub and Grub	43.679181	-79.297215	Nightlife Spot
3	East Toronto	The Beaches	43.676357	-79.293031	Upper Beaches	43.680563	-79.292869	Outdoors & Recreation
4	East Toronto	The Danforth West,Riverdale	43.679557	-79.352188	Pantheon	43.677621	-79.351434	Food

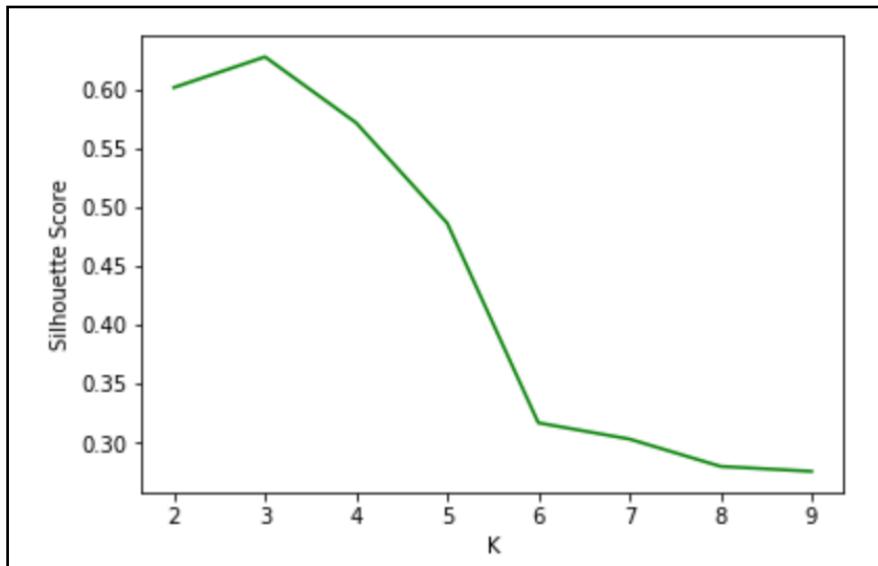
Then, using "one hot encoding", we can list all the Categories of the closest 100 venues for each neighborhood. Finally, we use “groupby” to get the aggregate data, which are the features that we will use for clustering.

Cluster Labels	Postcode	Borough	Neighbourhood	Latitude	Longitude	Arts & Entertainment	College & University	Event	Food	N
0	None	M4E	East Toronto	The Beaches	43.676357	-79.293031	0.000000	0.0	0.0	0.000000 0.
1	None	M4K	East Toronto	The Danforth West,Riverdale	43.679557	-79.352188	0.000000	0.0	0.0	0.681818 0.
2	None	M4L	East Toronto	The Beaches West,India Bazaar	43.668999	-79.315572	0.052632	0.0	0.0	0.578947 0.
3	None	M4M	East Toronto	Studio District	43.659526	-79.340923	0.000000	0.0	0.0	0.605263 0.
4	None	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790	0.000000	0.0	0.0	0.000000 0.

### 3.2.1.2 Determine the number of clusters for KMeans clustering

There are several metrics to evaluate a good KMeans clustering, such as: SSE(sum of the squared errors), Silhouette Coefficient. In this report, we will use Silhouette Coefficient to find the best K. Here

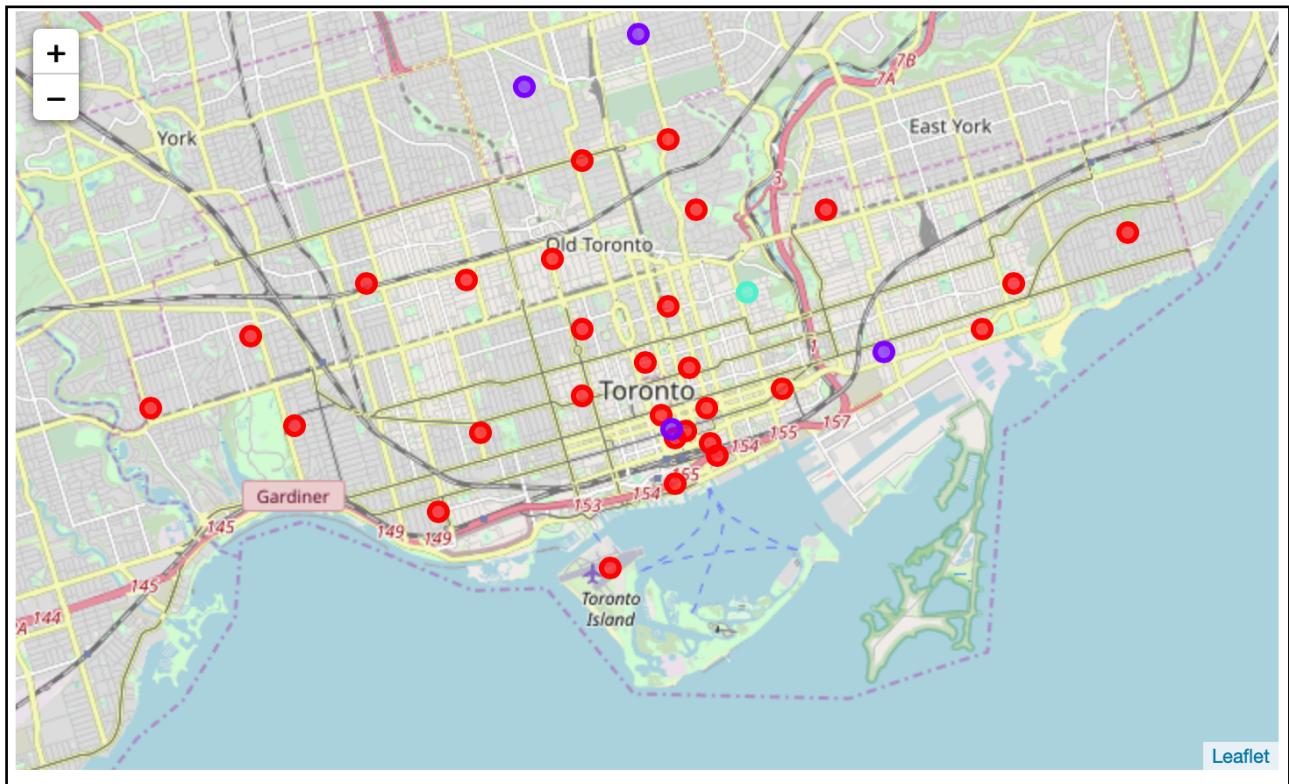
below is the plot of Silhouette Coefficient variance with K:



When K=3, the score is more than 0.5, which is pretty good for the clustering. We will therefore cluster the data of neighborhood into 3 clusters.

### 3.2.1.3 KMeans clustering

Here below is the cluster results showing in the map.



We can examine each cluster and determine the discriminating venue categories that distinguish each cluster.

First 5 rows of cluster 0:

Borough	Neighborhoods	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
Central Toronto	Davisville	Food	Shop & Service	Outdoors & Recreation	Nightlife Spot	Travel & Transport	Professional & Other Places	Event	College & University
Central Toronto	Davisville North	Shop & Service	Outdoors & Recreation	Food	Travel & Transport	Professional & Other Places	Nightlife Spot	Event	College & University
Central Toronto	Deer Park, Forest Hill, Rathnelly, South Hill, Summerhill West	Food	Nightlife Spot	Shop & Service	Travel & Transport	Professional & Other Places	Outdoors & Recreation	Event	College & University
Central Toronto	North Toronto West	Food	Shop & Service	Outdoors & Recreation	Travel & Transport	Professional & Other Places	Nightlife Spot	Event	College & University
Central Toronto	The Annex, North Midtown, Yorkville	Food	Shop & Service	Travel & Transport	Outdoors & Recreation	Nightlife Spot	Arts & Entertainment	Professional & Other Places	Event

Cluster 1:

Borough	Neighborhoods	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
Central Toronto	Forest Hill North, Forest Hill West	Outdoors & Recreation	Shop & Service	Food	Travel & Transport	Professional & Other Places	Nightlife Spot	Event	College & University
Central Toronto	Moore Park, Summerhill East	Outdoors & Recreation	Travel & Transport	Shop & Service	Professional & Other Places	Nightlife Spot	Food	Event	College & University
Central Toronto	Roselawn	Shop & Service	Outdoors & Recreation	Travel & Transport	Professional & Other Places	Nightlife Spot	Food	Event	College & University
Downtown Toronto	Rosedale	Outdoors & Recreation	Professional & Other Places	Travel & Transport	Shop & Service	Nightlife Spot	Food	Event	College & University
East Toronto	The Beaches	Outdoors & Recreation	Shop & Service	Nightlife Spot	Travel & Transport	Professional & Other Places	Food	Event	College & University

Cluster 2:

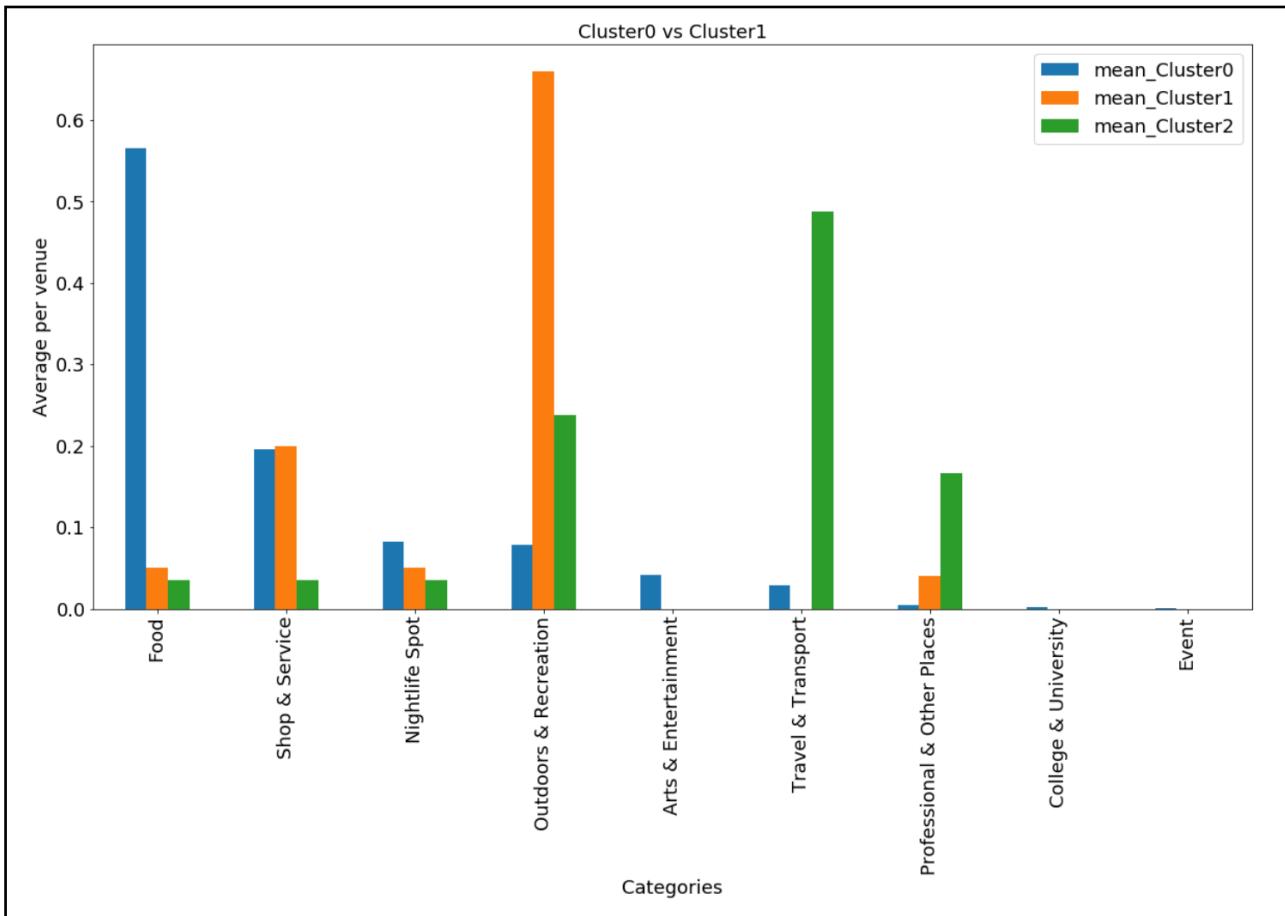
Borough	Neighborhoods	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
Central Toronto	Lawrence Park	Travel & Transport	Professional & Other Places	Outdoors & Recreation	Shop & Service	Nightlife Spot	Food	Event	College & University
Downtown Toronto	CN Tower, Bathurst Quay, Island airport, Harbourfront West, King and Spadina, Railway Lands, South Niagara	Travel & Transport	Outdoors & Recreation	Shop & Service	Nightlife Spot	Food	Professional & Other Places	Event	College & University

### 3.2.2 Conclusion of Venues' feature based on first level categories

Merge information of each cluster table and the table of feature for clustering in 3.2.1.1, we can get aggregate data of average number category per venue for each cluster as below:

	Arts & Entertainment	College & University	Event	Food	Nightlife Spot	Outdoors & Recreation	Professional & Other Places	Shop & Service	Travel & Transport
mean_Cluster0	0.041283	0.00222	0.000323	0.565401	0.082526	0.078470	0.005124	0.195561	0.029093
mean_Cluster1	0.000000	0.000000	0.000000	0.050000	0.050000	0.660000	0.040000	0.200000	0.000000
mean_Cluster2	0.000000	0.000000	0.000000	0.035714	0.035714	0.238095	0.166667	0.035714	0.488095

Based on this data, we can plot a bar chart for visualization:



Conclusion:

- **Cluster 0:** The average of Food and shop venues are much higher than other two clusters, so these are **food and shop** concentration area.
- **Cluster 1:** The average of Outdoor and Recreation are much higher than other two clusters venues, so these are **Outdoor and Recreation** concentration area.
- **Cluster 2:** The average of Travel and Transport are much higher than other two clusters venues, so these are **Travel and Transport** concentration area.

We can see from the table of most common venues for Cluster 1 and Cluster 2, that there is only one food spot for each cluster in the top 100 venues. So there's not much competition to be considered. Therefore, we only look at Cluster 0 in the remainder of our analysis of the competition.

**Note:** One could think that Clusters 1 and 2 would be better choices to open a restaurant, because there is very little competition. However, we do not want to discard Cluster 0 yet. Indeed, the fact that Clusters 1 and 2 have very little F&B presence is suspicious.

### 3.2.3 Cluster Neighborhoods based on the second level of venues' categories

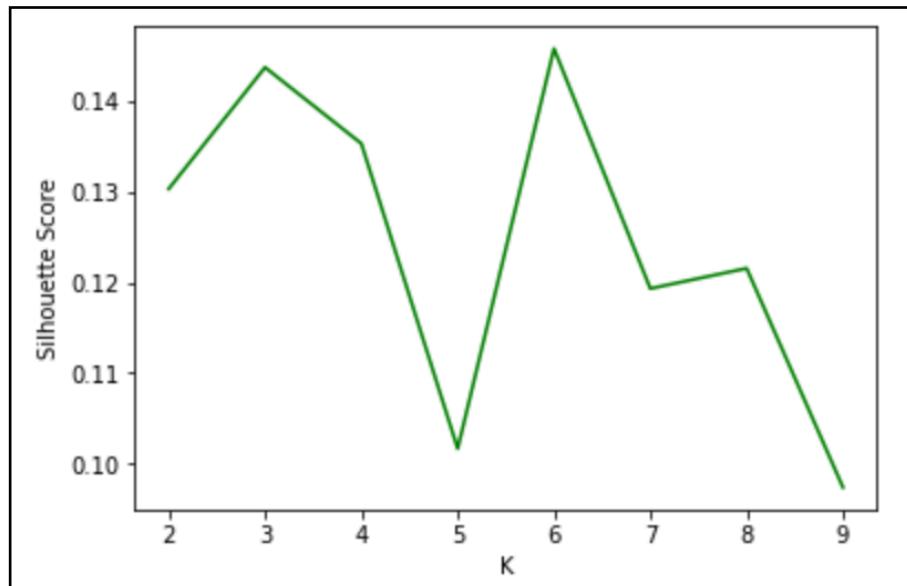
#### 3.2.3.1 Secondary Clustering of Neighborhoods in Cluster 0

In this section we want to analyze the competition, and because the clustering process is similar as the clustering based on first category of 3.2.1, we can already prepare a feature table. That is to say, we will only keep food and beverage venues, and be specific about the particular type of restaurant/cafe they are running. We will only consider the neighborhood in Cluster 0 for this section.

Here shows the first 5 rows of the features for secondary clustering:

Borough	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue new Category
4	East Toronto	The Danforth West,Riverdale	43.679557	-79.352188	Pantheon	43.677621	-79.351434
5	East Toronto	The Danforth West,Riverdale	43.679557	-79.352188	Dolce Gelato	43.677773	-79.351187
7	East Toronto	The Danforth West,Riverdale	43.679557	-79.352188	Mezes	43.677962	-79.350196
8	East Toronto	The Danforth West,Riverdale	43.679557	-79.352188	La Diperie	43.677530	-79.352295
9	East Toronto	The Danforth West,Riverdale	43.679557	-79.352188	Cafe Fiorentina	43.677743	-79.350115

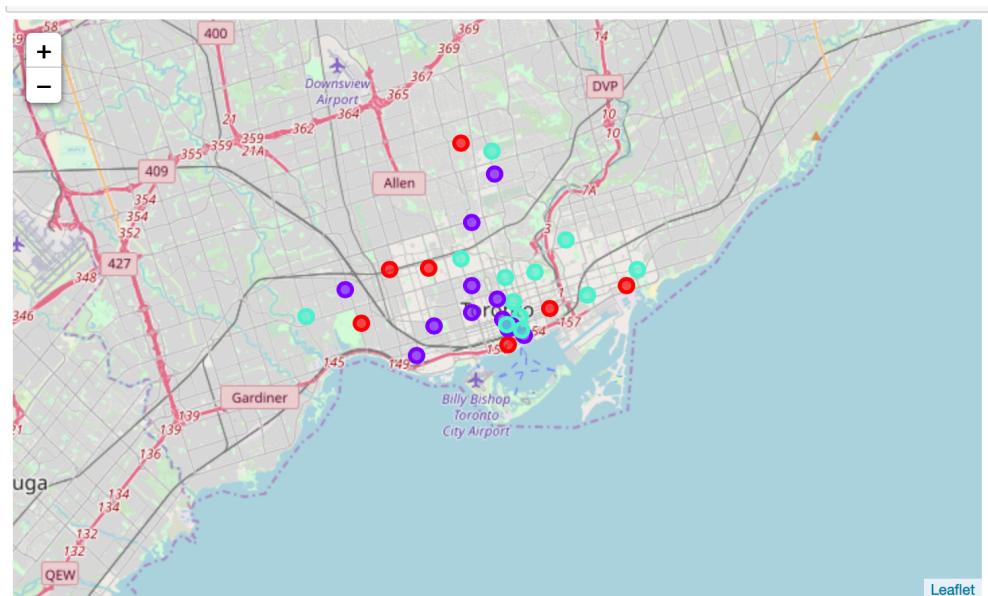
We determine the number of clusters by silhouette coefficient, which is the same metrics as in 3.2.1.



Even though,  $k=6$  has the highest silhouette coefficient, there are only 30 neighborhoods to be clustered. Six cluster seems unnecessarily high, we don't want to cluster the data into too small pieces. The second highest silhouette score is when  $k=3$ , this seems reasonable for our case.

Thus we will use KMeans to cluster the

### 3.2.3.2 Visualize the Cluster result



Cluster 0.0:

Borough	Neighborhoods	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Central Toronto	The Annex, North Midtown, Yorkville	Vegetarian / Vegan Restaurant	Asian Restaurant	Café	Coffee Shop	Italian Restaurant	Bakery	Restaurant	Pizza Place	Dessert Shop	Indian Restaurant
Downtown Toronto	Chinatown, Greektown, Park, Kensington Market	Asian Restaurant	Café	Vegetarian / Vegan Restaurant	Dessert Shop	Mexican Restaurant	Bakery	Coffee Shop	Burger Joint	French Restaurant	Comfort Food Restaurant
Downtown Toronto	Harbord, University of Toronto	Café	Asian Restaurant	Bakery	Dessert Shop	Vegetarian / Vegan Restaurant	Restaurant	Mexican Restaurant	Pizza Place	Comfort Food Restaurant	Bubble Tea Shop
East Toronto	The Beaches, West, India Bazaar	Indian Restaurant	Café	Coffee Shop	Asian Restaurant	Italian Restaurant	Mexican Restaurant	Restaurant	Sandwich Place	Fast Food Restaurant	Bakery
East Toronto	The Danforth, West, Riverdale	Greek Restaurant	Coffee Shop	Café	Asian Restaurant	Pizza Place	Dessert Shop	Italian Restaurant	Sandwich Place	Caribbean Restaurant	Portuguese Restaurant
West Toronto	Little Portugal, Trinity	Asian Restaurant	Café	Coffee Shop	Restaurant	Bakery	Pizza Place	Vegetarian / Vegan Restaurant	Italian Restaurant	Dessert Shop	Spanish Restaurant
West Toronto	Ranneymede, Swansea	Coffee Shop	Café	Bakery	Asian Restaurant	Pizza Place	Italian Restaurant	Dessert Shop	Mexican Restaurant	Falafel Restaurant	Gastropub

Cluster 0.1:

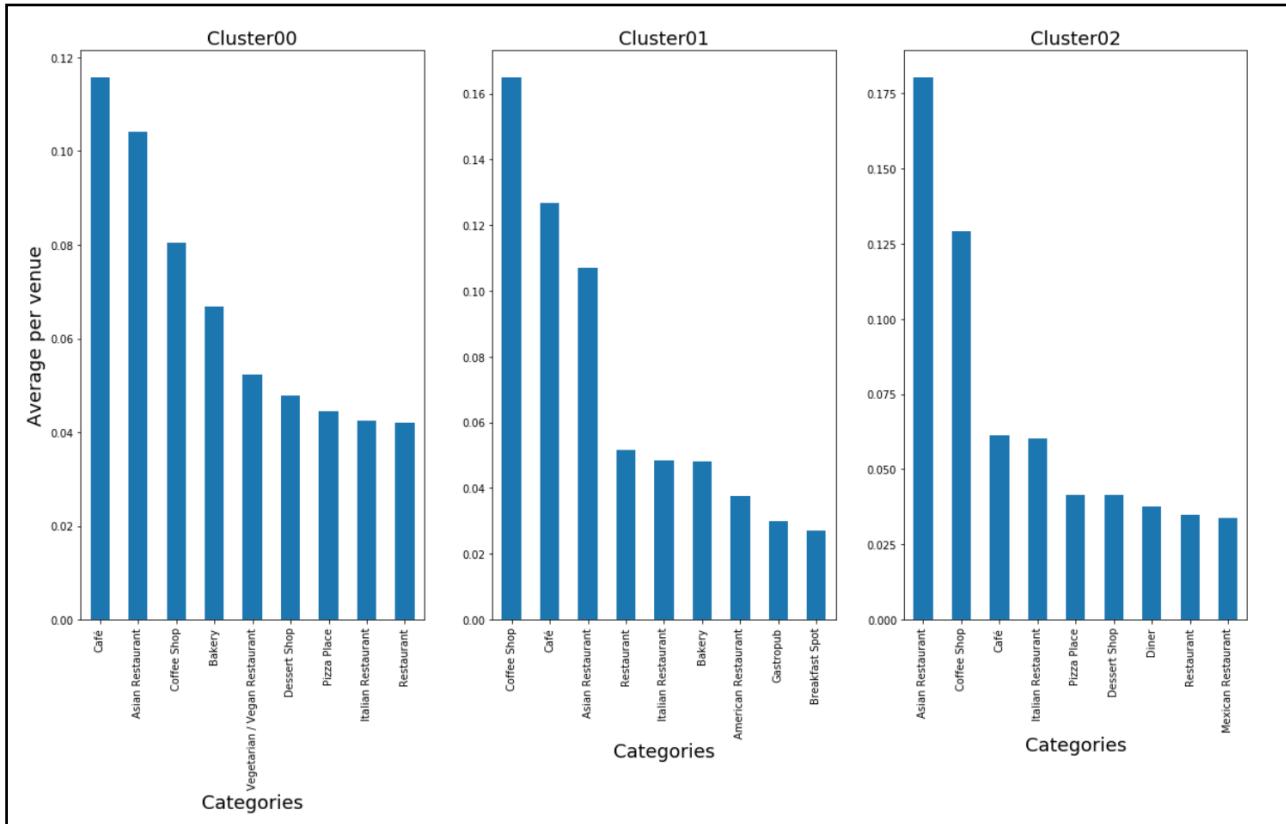
Borough	Neighborhoods	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Downtown Toronto	Adelaide,Kin g,Richmond	Asian Restaurant	Café	Coffee Shop	American Restaurant	Pizza Place	Greek Restaurant	Latin American Restaurant	Gastropub	Vegetarian / Vegan Restaurant	Mexican Restaurant
Downtown Toronto	Berczy Park	Coffee Shop	Café	Asian Restaurant	Restaurant	Bakery	Seafood Restaurant	Gastropub	Breakfast Spot	Italian Restaurant	Steakhouse
Downtown Toronto	Commerce Court,Victori a Hotel	Coffee Shop	Asian Restaurant	Café	American Restaurant	Gastropub	Steakhouse	Restaurant	Bakery	Vegetarian / Vegan Restaurant	Deli / Bodega
Downtown Toronto	Design Exchange,T oronto Dominion Centre	Coffee Shop	Café	Asian Restaurant	American Restaurant	Italian Restaurant	Steakhouse	Restaurant	Gastropub	Vegetarian / Vegan Restaurant	Deli / Bodega
Downtown Toronto	First Canadian Place,Under ground city	Coffee Shop	Asian Restaurant	Café	American Restaurant	Italian Restaurant	Steakhouse	Bakery	Restaurant	Gastropub	Latin American Restaurant
Downtown Toronto	Harbourfront East,Toronto Islands,Unio n Station	Coffee Shop	Asian Restaurant	Café	Italian Restaurant	Restaurant	Deli / Bodega	Steakhouse	Bakery	Fried Chicken Joint	Dessert Shop
Downtown Toronto	Harbourfront ,Regent Park	Coffee Shop	Café	Asian Restaurant	Italian Restaurant	Breakfast Spot	Diner	Restaurant	Bakery	Mexican Restaurant	Pizza Place
Downtown Toronto	St. James Town	Coffee Shop	Café	Restaurant	American Restaurant	Asian Restaurant	Bakery	Italian Restaurant	Seafood Restaurant	Gastropub	Breakfast Spot
Downtown Toronto	Stn A PO Boxes 25 The Esplanade	Coffee Shop	Café	Asian Restaurant	Restaurant	Bakery	Gastropub	Seafood Restaurant	Breakfast Spot	American Restaurant	Italian Restaurant
West Toronto	Brockton,Ex hibition Place,Parkdale Village	Café	Coffee Shop	Asian Restaurant	Bakery	Restaurant	Sandwich Place	Italian Restaurant	Vegetarian / Vegan Restaurant	Indian Restaurant	Caribbean Restaurant
West Toronto	Dovercourt Village,Duffe rin	Café	Coffee Shop	Asian Restaurant	Portuguese Restaurant	Italian Restaurant	Bakery	Mediterrane an Restaurant	Mexican Restaurant	Middle Eastern Restaurant	Juice Bar
West Toronto	High Park,The Junction South	Asian Restaurant	Café	Coffee Shop	Italian Restaurant	Deli / Bodega	Gastropub	Food Truck	Breakfast Spot	Bakery	Dessert Shop

### Cluster 0.2:

Borough	Neighborhoods	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Central Toronto	Davisville	Asian Restaurant	Coffee Shop	Italian Restaurant	Dessert Shop	Indian Restaurant	Restaurant	Mexican Restaurant	Gastropub	Café	Pizza Place
Central Toronto	Davisville North	Coffee Shop	Asian Restaurant	Dessert Shop	Fast Food Restaurant	Italian Restaurant	Café	Pizza Place	Mexican Restaurant	Diner	Restaurant
Central Toronto	Deer Park,Forest Hill SE,Rathnelly,South Hill,Summerhill West	Asian Restaurant	Coffee Shop	Italian Restaurant	Pizza Place	Restaurant	Café	Bagel Shop	Sandwich Place	Bistro	German Restaurant
Central Toronto	North Toronto West	Asian Restaurant	Coffee Shop	Italian Restaurant	Café	Dessert Shop	Diner	Mexican Restaurant	Bakery	Fast Food Restaurant	Tea Room
Downtown Toronto	Cabbagetown,St. James Town	Asian Restaurant	Diner	Coffee Shop	Gastropub	Café	Bistro	Indian Restaurant	American Restaurant	Steakhouse	Bakery
Downtown Toronto	Central Bay Street	Asian Restaurant	Coffee Shop	Café	Dessert Shop	Vegetarian / Vegan Restaurant	Italian Restaurant	Mexican Restaurant	Diner	Tea Room	Gastropub
Downtown Toronto	Christie	Asian Restaurant	Café	Coffee Shop	Dessert Shop	Mexican Restaurant	Indian Restaurant	Latin American Restaurant	Pizza Place	Sandwich Place	Vegetarian / Vegan Restaurant
Downtown Toronto	Church and Wellesley	Asian Restaurant	Coffee Shop	Café	Mediterrane an Restaurant	Mexican Restaurant	Middle Eastern Restaurant	Bubble Tea Shop	Burger Joint	Gastropub	Pizza Place
Downtown Toronto	Ryerson,Garden District	Asian Restaurant	Coffee Shop	Middle Eastern Restaurant	Diner	American Restaurant	Tea Room	Gastropub	Italian Restaurant	Café	Fast Food Restaurant
East Toronto	Business Reply Mail Processing Centre 969 Eastern	Coffee Shop	Asian Restaurant	Pizza Place	Italian Restaurant	Mexican Restaurant	Restaurant	Dessert Shop	Diner	Burger Joint	Breakfast Spot
East Toronto	Studio District	Asian Restaurant	Coffee Shop	American Restaurant	Bakery	Italian Restaurant	Café	Diner	French Restaurant	Gastropub	Sandwich Place
West Toronto	Parkdale,Roncesvall es	Asian Restaurant	Coffee Shop	Café	Pizza Place	Breakfast Spot	Restaurant	Eastern European Restaurant	Bakery	Sandwich Place	Dessert Shop

### 3.2.3.3 Comparison of Clusters

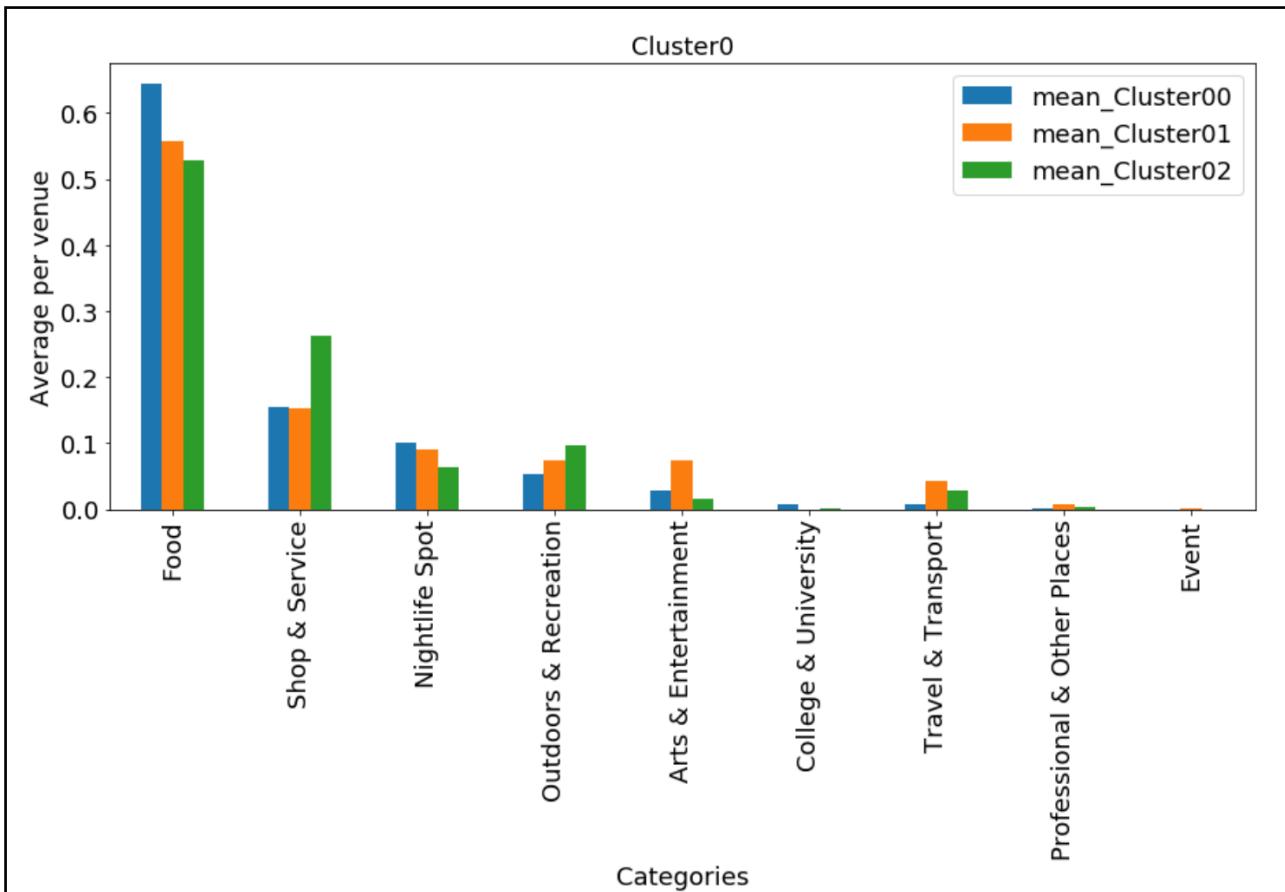
First, let's look at the distribution of food and beverage businesses for the 3 clusters.



Then, let's look at the other categories of venues for these 3 clusters, in order to have an idea of the neighborhood environment besides food industry.

Here's the aggregate data:

	Arts & Entertainment	College & University	Event	Food	Nightlife Spot	Outdoors & Recreation	Professional & Other Places	Shop & Service	Travel & Transport
<b>mean_Cluster00</b>	0.029266	0.008403	0.000000	0.644425	0.100969	0.054103	0.001429	0.154023	0.007381
<b>mean_Cluster01</b>	0.074302	0.000000	0.000833	0.557057	0.090802	0.073655	0.008384	0.152921	0.042045
<b>mean_Cluster02</b>	0.015273	0.000833	0.000000	0.527648	0.063490	0.097500	0.004018	0.262432	0.028806



Finally, our customer is interested in opening a café and/or Italian restaurant. Let's have a look at the distribution of these two type of business in these clusters.

	Café	Italian Restaurant
<b>mean_Cluster00</b>	0.115670	0.042510
<b>mean_Cluster01</b>	0.126843	0.048447
<b>mean_Cluster02</b>	0.061447	0.060102

### 3.2.3.4 Conclusion of Venues' features

- **Cluster 00 :**

Environment: The average of food & beverage business per venue is the highest among the three clusters. These are no doubt popular places for eating.

Competition: Café is the most common venue for this cluster, so lots of competition. However Italian restaurant is less common than the other two clusters. It should be interesting to look into it for opening of an **Italian Restaurant**.

- **Cluster 01 :**

Environment: The average of Arts & Entertainment business per venue is the highest among the three clusters. Therefore, the flow of customers should be good as well.

Competition: Even though the food & beverage business is less common than Cluster 00, but it's also more than 50%, which is not bad. But there are more Café and Italian restaurant per venue than Cluster 00, so more competition, it's less interesting than Cluster 00.

- **Cluster 02:**

Environment: Much less restaurant and café than Cluster 01 and 00, but more shopping places and more Outdoors & Recreation areas, thus the flow of customers should be good.

Competition: The density of café is half of cluster 01. It should be interesting for investment of a café given the relatively less competition and good flow.

## 4. Conclusion

Based on the above analysis, here are the lists for interesting locations of investment for a café or an Italian Restaurant.

List of neighborhoods for Italian Restaurant:

Borough	Neighbourhood	Latitude	Longitude
4 Central Toronto	North Toronto West	43.715383	-79.405678
9 Downtown Toronto	Harbourfront,Regent Park	43.654260	-79.360636
15 Downtown Toronto	Harbourfront East,Toronto Islands,Union Station	43.640816	-79.381752
23 Downtown Toronto	Christie	43.669542	-79.422564
24 West Toronto	Dovercourt Village,Dufferin	43.669005	-79.442259
28 West Toronto	Parkdale,Roncesvalles	43.648960	-79.456325
30 East Toronto	Business Reply Mail Processing Centre 969 Eastern	43.662744	-79.321558

List of neighborhoods for café:

Borough	Neighbourhood	Latitude	Longitude
0 East Toronto	The Danforth West,Riverdale	43.679557	-79.352188
1 East Toronto	The Beaches West,India Bazaar	43.668999	-79.315572
2 East Toronto	Studio District	43.659526	-79.340923
3 Central Toronto	Davisville North	43.712751	-79.390197
7 Downtown Toronto	Cabbagetown,St. James Town	43.667967	-79.367675
8 Downtown Toronto	Church and Wellesley	43.665860	-79.383160
10 Downtown Toronto	Ryerson,Garden District	43.657162	-79.378937
11 Downtown Toronto	St. James Town	43.651494	-79.375418
18 Central Toronto	The Annex,North Midtown,Yorkville	43.672710	-79.405678
21 Downtown Toronto	Stn A PO Boxes 25 The Esplanade	43.646435	-79.374846
22 Downtown Toronto	First Canadian Place,Underground city	43.648429	-79.382280
29 West Toronto	Runnymede,Swansea	43.651571	-79.484450

Visualization on a map:

